

Figure 1 consists of 12 histograms arranged in a single row. Each histogram represents the distribution of the number of non-zero elements in the vector x for a specific value of n . The x-axis for all histograms is labeled 'Number of non-zero elements' and ranges from 0 to 120. The y-axis is labeled 'Frequency' and ranges from 0 to 100. The histograms are labeled with n values: 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, and 120. For $n=10$, the distribution is centered around 60. As n increases, the distribution shifts to the right, and the peak frequency decreases. For example, for $n=120$, the distribution is centered around 120, and the frequency is very low.

Inventor(s) : Naganori NUMAO

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is based upon and claims priority of Japanese Patent Application No. 2000-206129, filed on July 7, 2000, the contents being incorporated herein by reference.

【発明の背景】 (Background of the Invention)

【発明の分野】 (Field of the Invention)

本発明は任意の蛋白質（又は核酸配列）の新しい生物・機能活性又は結合相手を効率良く予測するための知的情報技術に関する。

【関連技術の説明】 (Description of the Related Art)

蛋白質（又は核酸配列）の生物・機能活性の予測法は、アミノ酸配列（又は核酸配列）の相同性に依存した多数の方法が開発されているが、それらの方法は個別的であり一般性に欠けている。従って、比較する配列間の相同性が低い場合には生物・機能活性の予測は極めて困難である。現在、簡便で一般性の高い方法の開発が社会から強力に要請されている。特に、現在ゲノム解析により得られる蛋白質（又は核酸配列）の機能解明に関する新しい方法論の提供が強く望まれている。

現在、任意の蛋白質（または核酸配列）の生物・機能活性を予測する場合、最もよく用いられる方法はその蛋白質の全アミノ酸配列（又は核酸配列）中に存在するモチーフの探索である（A. Bairoch et al., Nucleic Acids Res., 20, 2019-2022 (1992)）。この方法は、任意の蛋白質の機能を効率良く予測する方法の1つとして知られている（M. J. E. Sternberg, CABIOS, 7, 257-260 (1991)）。また、もう1つ別の検索方法として、任意の新規蛋白質の全アミノ酸配列と生物活性既知である蛋白質の全アミノ酸配列とのホモロジー検索法である（R. F. Doolittle et al, Nature 307, 558-560 (1984)）。これまでのところ、それら2つの方法が最も良く使われる機能予測法であるが、個別的であり、一般性に欠ける。即ち、比較する2つの蛋白質のアミノ酸配列にモチーフや相同性が見出せない場合には生物・機能活性を予測できない。たとえそれらのアミノ酸配列の相同性が高くとも生物活性

が同じであるとは限らないことも良く知られている。依然としてそれらの方法を改良が望まれている。

1985年に、ヴェルコヴィッチ等は生物活性が全く同じである少なくとも2つ以上の蛋白質の全アミノ酸配列（又は核酸配列）のアミノ酸（又は核酸残基）にEIIP指標数を付与し、それらをデジタルプロセッシング法で計算するとそれらのアミノ酸配列や核酸配列の相同性には関係なくそれらの蛋白質の振動数値は特定の値に収斂する画期的な方法論を報告した（V. Veljkovic et al, IEEE 32, 337-3418(1985) ; V. Veljkovic et al, Cancer Biochem. Biophys., 9, 139-148 (1987)）。この方法は、同じ生物活性を有する有機高分子化合物（蛋白質類や核酸配列）の分類法の1つとして非常に有用かもしれない。彼等はこれまで蛋白質の特異的振動数値について多数報告（V. Veljkovic et al, Cancer Biochem. Biophys., 9, 139-148 (1987) ; I. Cosic, IEEE, 41, 1101-1114 (1994)）しているが、蛋白質単独での機能・活性予測や特異的な分子間相互作用を全く言及していない。それらの課題を解決するためには、任意のアミノ酸配列（又は核酸配列）の生物・機能活性部位に由来する特異的周波数スペクトルを新規方法で抽出しなければならない。

本発明者は、既に蛋白質の活性部位（基質結合部位や触媒活性部位）に関する一般的予測法を報告した（N. Numao et al, Biol. Pharm. Bull., 16, 1160-1163 (1993)）。即ち、任意の蛋白質の活性部位領域に、13種類の相補的アミノ酸ユニット〔GT、AS、GA、ID、TR、SR、LK、TXW、VXH、MXH、WXP、AXC、GXS〔（G、T、A、S、I、D、R、L、K、W、V、H、M、P、C、Xはグリシン、スレオニン、アラニン、セリン、イソロイシン、アスパラギン酸、アルギニン、ロイシン、リジン、トリプトファン、バリン、ヒスチジン、メチオニン、プロリン、システイン及び20種類のアミノ酸の何れかを意味する）〕又はそれらの逆配列の相補的アミノ酸ユニットが少なくとも1つ以上存在することが報告されている。更に、本発明者らは蛋白質の活性部位予測法をリボザイムなどの核酸配列の触媒活性領域を予測する場合、翻訳した仮想アミノ酸配列中に上記のモチーフ配列の存在するところが活性領域であることを報告している（N. Numao et al., EP Appl. No 91311129.0 (1991)）。その方法はあらゆる種類の蛋白質のアミノ酸配列や核酸配列の活性部位を予測する方法として有用であるが、生物・機能活性の種類と活性部位領域との関連性や特異的な分子間相互作用を全く言及していない。それらの課題を解決するためには、分子間相互作用に参与するアミノ酸配列（又は核酸配列）領域に新規方法で物理定数を与え、数理的に解析しなければ

ならない。

【発明の要約】 (Summary of the Invention)

本発明の目的は、従来法より効率よく、より一般的に蛋白質（または核酸配列）の活性部位由来の生物・機能活性を予測できる方法を提供することである。

具体的には、天然型、非天然型由来の、1) 任意のアミノ酸配列のアミノ酸残基に、EIIP (Electron-ion interaction potential) 指標数を付与し、得られた数値列（以下EIIP列と称す）を離散フーリエ変換（以下DFTと称す）して得られる全アミノ酸配列周波数スペクトル及び/又は2) その任意のアミノ酸配列に存在し、活性部位に関与する既知モチーフを1つ以上含有する2～64個のアミノ酸残基を含有するアミノ酸配列領域のアミノ酸にEIIP指標数を付与し、得られたEIIP列をDFTして得られる周波数スペクトル（以下、活性部位周波数スペクトルと称す）及び/又は3) その任意のアミノ酸配列のアミノ酸配列に学術的に対応する核酸配列の核酸残基にEIIP指標数を付与し、得られたEIIP列をDFTして得られる全核酸配列周波数スペクトル及び/又は4) 任意の一本鎖の核酸配列の核酸残基にEIIP指標数を付与し、得られたEIIP列をDFTして得られる全核酸配列周波数スペクトル及び/又は5) 任意の一本鎖の核酸配列に水素結合する核酸配列の核酸残基にEIIP指標数を付与し、得られたEIIP列をDFTして得られる全核酸配列周波数スペクトルを用いて、所望の任意のアミノ酸配列および又は核酸配列の生物・機能活性（又は結合活性）を予測する方法を提供する。プリオン蛋白質と β アミロイド蛋白質前駆体のオキザロアセテートに対する脱炭酸活性や、カルシトニンとヒト成長ホルモンとの類似の生物活性や、エボラウィルスの55kdTNFレセプターへの結合等、これまで知られていない新しい生物・機能活性が簡単に予測することを目的とする。

本発明の任意の蛋白質の生物・機能活性及び/又は結合活性を予測する方法は、天然型、非天然型由来の任意のアミノ酸配列のアミノ酸残基に、EIIP (Electron-ion interaction potential) 指標数を付与し、得られたEIIP列をDFTして得られた全アミノ酸配列周波数スペクトルと、前記任意のアミノ酸配列に存在し、活性部位に関与する既知モチーフを1つ以上含有する2～64個のアミノ酸残基を含有するアミノ酸配列領域のアミノ酸にEIIP

指標数を付与し、得られたEIIP列をDFTして得られた活性部位周波数スペクトルとを求め、前記全アミノ酸配列周波数スペクトルと前記活性部位周波数スペクトルとをクロスしたスペクトルから、その蛋白質の活性部位由来の特異的振動数値を選択し、特異的振動数値の近似値を検索することにより任意の蛋白質の生物・機能活性及び/又は結合活性を予測する方法である。

本発明の前記方法の一態様では、活性部位の信号としての既知モチーフとして、GT、AS、GA、ID、TR、SR、LK、TXW、VXH、MXH、WXP、AXC、GXS (G、T、A、S、I、D、R、L、K、W、V、H、M、P、C、Xはグリシン、スレオニン、アラニン、セリン、イソロイシン、アスパラギン酸、アルギニン、ロイシン、リジン、トリプトファン、バリン、ヒスチジン、メチオニン、プロリン、システイン及び20種類のアミノ酸の何れかを意味する) 及び/又はそれらの逆配列の何れか1つ以上を用いる。

本発明の任意の蛋白質の生物・機能活性及び/又は結合活性を予測する方法は、天然型、非天然型由来の任意のアミノ酸配列のアミノ酸残基に、EIIP (Electron-ion interaction potential) 指標数を付与し、得られたEIIP列をDFTして得られた全アミノ酸配列周波数スペクトルと、前記アミノ酸配列に学術的に対応する核酸配列領域の核酸残基にEIIP指標数を付与し、得られたEIIP列をDFTして得られた全核酸配列周波数スペクトルとを求め、前記全アミノ酸配列周波数スペクトルと前記全核酸配列周波数スペクトルとをクロスしたスペクトルからその蛋白質の活性部位由来の特異的振動数値を選択し、特異的振動数値の近似値を検索することにより任意の蛋白質の生物・機能活性及び/又は結合活性を予測する方法である。

本発明の任意の核酸配列の生物・機能活性及び/又は結合活性を予測する方法は、天然型、非天然型由来の任意の一本鎖の核酸配列の核酸残基にEIIP指標数を付与し、得られたEIIP列をDFTして得られた第1の全核酸配列周波数スペクトルと、前記核酸配列に水素結合によって結合する核酸配列の核酸残基にEIIP指標数を付与し、得られたEIIP列をDFTして得られた第2の全核酸配列周波数スペクトルとを求め、前記第1の全核酸配列周波数スペクトルと前記第2の全核酸配列周波数スペクトルとをクロスしたスペクトルからその核酸配列の特異的振動数値を選択し、特異的振動数値の近似値を検索することにより任意の核酸配列の生物・機能活性及び/又は結合活性を予測する方法である。

また、本発明は、上記の各スペクトルを組み合わせることによって、天然型、非天然型由来の任意のアミノ酸配列と別の核酸配列の生物・機能活性及び/又は結合活性を予測する方法である。

また、本発明は、上記の各スペクトルを組み合わせることによって、天然型、非天然型由来の任意のアミノ酸配列又は核酸配列の活性部位を予測する方法である。

【図面の簡単な説明】 (Brief Description of the Drawings)

図1 A～1 Gは、本発明の演算方法例を示す図である。

図2は、マゲイニン2前駆体の自己クロススペクトルを示す図である。

図3は、マゲイニン2前駆体とマゲイニン2のクロススペクトルを示す図である。

図4は、マゲイニン2前駆体と221-233をロイシンで置換したマゲイニン2前駆体のクロススペクトルを示す図である。

図5は、マゲイニン2前駆体の混合周波数スペクトルを示す図である。

図6は、マゲイニン2の混合周波数スペクトルを示す図である。

図7は、MSI-78Aの混合周波数スペクトルを示す図である。

図8は、鮭カルシトニン前駆体の自己クロススペクトルを示す図である。

図9は、鮭カルシトニン前駆体と鮭カルシトニンとのクロススペクトルを示す図である。

図10は、鮭カルシトニン前駆体と鮭カルシトニン前駆体の83-114をロイシンで置換したロイシンで置換した鮭カルシトニン前駆体のクロススペクトルを示す図である。

図11は、鮭カルシトニン前駆体と鮭カルシトニン前駆体の83-114以外領域をロイシンで置換した鮭カルシトニン前駆体のクロススペクトルを示す図である。

図12は、ガンマインターフェロンの自己クロススペクトルを示す図である。

図13は、ガンマインターフェロンとその活性部位領域(132～162)とのクロススペクトルを示す図である。

図14は、ガンマインターフェロンと活性部位領域(132～162)以外領域をロイシンで置換したガンマインターフェロンのクロススペクトルを示す図である。

図15は、ガンマインターフェロンの混合周波数スペクトルを示す図である。

図16は、ガンマインターフェロンレセプターの混合周波数スペクトルを示す図である。

図17は、Gal4pの混合周波数スペクトルを示す図である。

図18は、Gal4pの活性部位領域（14～57）の混合周波数スペクトルを示す図である。

図19は、Gal7のプロモーター領域の均一周波数スペクトルを示す図である。

図20は、ウロキナーゼの混合周波数スペクトルを示す図である。

図21は、サブチリシンの混合周波数スペクトルを示す図である。

図22は、プリオン蛋白質の自己クロススペクトルを示す図である。

図23は、プリオン蛋白質とプリオン蛋白質の109～131領域とのクロススペクトルを示す図である。

図24は、プリオン蛋白質とその109～131領域のアミノ酸残基全てをロイシンで置換したプリオン蛋白質とのクロススペクトルを示す図である。

図25は、プリオン蛋白質の109～131領域の混合周波数スペクトルを示す図である。

図26は、プリオン蛋白質の110～126領域の混合周波数スペクトルを示す図である。

図27は、アミロイド蛋白質前駆体の自己クロススペクトルを示す図である。

図28は、アミロイド蛋白質前駆体とその650～680領域とのクロススペクトルを示す図である。

図29は、アミロイド蛋白質前駆体とその650～680領域のアミノ酸残基全てをロイシンで置換したアミロイド蛋白質前駆体とのクロススペクトルを示す図である。

図30は、アミロイド蛋白質前駆体とその289～364領域とのクロススペクトルを示す図である。

図31は、ヒト成長ホルモンの自己クロススペクトルを示す図である。

図32は、ヒト成長ホルモンとその190～217領域のアミノ酸残基全てをロイシンで置換したヒト成長ホルモンとのクロススペクトルを示す図である。

図33は、ヒト成長ホルモンとその190～217領域以外のアミノ酸残基全てをロイシンで置換したヒト成長ホルモンとのクロススペクトルを示す図である。

図34は、ヒト成長ホルモンと鮭カルシトニン前駆体とのクロススペクトルを示す図である。

図35は、ヒト成長ホルモンと鮭カルシトニンとのクロススペクトルを示す図である。

図36は、エボラウィルスの混合周波数スペクトルを示す図である。

【好適実施例の詳細な説明】 (Detailed Description of the Preferred Embodiments)

本発明者は鋭意検討した結果、天然型又は非天然型のアミノ酸配列（又は核酸配列）とその配列中に存在する活性部位領域を組み合わせることによって、単一蛋白質や核酸配列の活性部位由来の生物・機能活性を予測しうる方法を見出した。この方法では、GenBank、EMBL、PIR、SWISS-PROT等のデータベース（図1Aに登録された図1Bの蛋白質（アミノ酸配列）又は核酸配列）を用いる。

即ち、第一段階はヴェルコヴィッチ等（V. Veljkovic et al, Cancer Biochem. Biophys., 9, 139-148 (1987); I. Cosic, IEEE, 41, 1101-1114 (1994)）の方法に従って、天然型又は非天然型の任意の蛋白質の全アミノ酸配列のアミノ酸（又は核酸）にEIIP (Electron-ion interaction potential) 指標数を付与し、図1Cの得られた数値列（以下EIIP列と称す）を離散フーリエ変換（以下DFTと称す）することによって全アミノ酸配列のEIIP列をDFTにより得られた図1Dの周波数スペクトル（以下全アミノ酸周波数スペクトルと称す）を自己クロスし、ピークの相対強度（高さ）に基づいて多数のピークから、その蛋白質分子全体に由来する非特異的振動数値を選択する。ピークを選択方法は振動数値の高い順に30ポイントを選ぶ。好ましくは3から12ポイントまでのうちの所定値を選ぶ。このピークを選択方法は以下の説明においても同様である。

ヴェルコヴィッチ等の方法を具体的に説明する。即ち、鎖長Lのアミノ酸配列 $a_0a_1a_2a_3 \cdots a_{L-3}a_{L-2}a_{L-1}$ の残基にEIIP指標値（V. Veljkovic et al, Cancer Biochem. Biophys., 9, 139-148 (1987)）を付与し、EIIP列 $f_n (n=0, 1, 2, 3, \cdots, L-3, L-2, L-1)$ （図1C）を作成する。但し、EIIP列は2のべき乗になるように、アミノ酸配列を延長する。延長された部分の数値列の値はアミノ酸配列の平均EIIP指標値を採用する。

アミノ酸残基に与えるEIIP指標値は次の通りである。

Leu 0.0000 Ile 0.0000 Asn 0.0036 Gly 0.0050 Val 0.0057 Glu 0.0058 Pro 0.0198

His 0.0242 Lys 0.0371 Ala 0.0373 Tyr 0.0516 Trp 0.0548 Gln 0.0761 Met 0.0823
Ser 0.0829 Cys 0.0829 Thr 0.0941 Phe 0.0946 Arg 0.0959 Asp 0.1263

得られたEIIP列 f_n ($n=0, 1, 2, 3, \dots, L-3, L-2, L-1$) (図1C)を、下記の離散フーリエ変換(DFT)式、即ち

$$F_m = \sum_{n=0,1,2,3,\dots,L-3,L-2,L-1} f_n \exp(2mn\pi i/L)$$

でもって処理し、周波数スペクトル F_m (図1D)を得る。 F_m は周期性条件を満たす。即ち、 $F_{m+m} = F_m^*$ である。この条件から、 $m = 0, 1, 2, \dots, L/2$ なる F_m のみが情報をもつ。

第二段階は、本発明者が既に報告 (N. Numao et al, Biol. Pharm. Bull., 16, 1160-1163 (1993)) した、任意の蛋白質の全アミノ酸配列中の活性部位に關与する既知モチーフとして、GT、AS、GA、ID、TR、SR、LK、TXW、VXH、MXH、WXP、AXC、GXS 又はその逆配列 (G、T、A、S、I、D、R、L、K、W、V、H、M、P、C、Xはグリシン、スレオニン、アラニン、セリン、イソロイシン、アスパラギン酸、アルギニン、ロイシン、リジン、トリプトファン、バリン、ヒスチジン、メチオニン、プロリン、システイン及び20種類のアミノ酸の何れかを意味する) のうち何れか1つ以上を含有する2~64個のアミノ酸残基からなる活性部位領域 (図1E) のアミノ酸 (又は核酸) にEIIP指標数を付与し、第一段階の方法と同様に、そのEIIP列をDFTすることによって得られた図1Fの周波数スペクトル (以下活性部位周波数スペクトルと称す) 及びその全アミノ酸配列のEIIP列をDFTにより得られた図1Dの全アミノ酸周波数スペクトルとクロス (掛け算) し、得られたクロススペクトル (図1G) のピークの相対強度に基づいて、多数のピークからその蛋白質の活性部位由来の振動数値を選択する。

第三段階は、上記の2~64個のアミノ酸残基からなる活性部位領域のアミノ酸残基全てを20種類のアミノ酸のうち何れか1つのアミノ酸を使って置換するか又は2~64個のアミノ酸残基からなる活性部位を変えず、それ以外の領域全てを20種類のアミノ酸のうち何れか1つのアミノ酸を使って置換した全アミノ酸配列 (以下、置換全アミノ酸配列と称する) のアミノ酸残基にEIIP指標数を付与し、得られたEIIP列をDFTすることによって置換全アミノ酸周波数スペクトルを求め、そのスペクトルをもとの全アミノ酸周波数スペクトルとクロスする。そして、第一段階と第二段階の方法で選択された活性部位由来の振動数値を第三段階の結果から確認する。

上記3段階方法を用いて、効率よく任意の蛋白質の活性部位由来の特異的振動数値を選

択し、その蛋白質の生物・機能活性を予測する方法を見出し、本発明を完成するに到った。

本発明者は、その発明過程を明らかにするために、23アミノ酸残基からなる抗菌活性を有するマゲイニン2を含有するマゲイニン前駆体を例示する。マゲイニン前駆体は303個のアミノ酸残基からなり、その配列中にはマゲイニン2が5つとマゲイニン1が1つコードされている (M. Zasloff, Proc. Natl. Acad. Sci., U.S.A., 84, 5449-5453 (1987))。それらのアミノ酸配列はSWISS-PROT に、核酸配列はGenBank に登録されている。

本発明者は問題を解決するための手段を開発するための作業仮説として、マゲイニン2 (又はマゲイニン1) のアミノ酸配列を前駆体アミノ酸配列に存在する活性部位領域と仮定した。

先ず、第一段階はヴェルコヴィッチ等の方法 (V. Veljkovic et al, IEEE 32, 337-3418(1985) ; V. Veljkovic et al, Cancer Biochem. Biophys., 9, 139-148 (1987)) に従って、マゲイニン前駆体の全アミノ酸周波数スペクトルを自己クロスすることにより、前駆体の全アミノ酸周波数スペクトル (図2) を求める。図2より、ピークのS/N (Sはシグナルピーク、Nはノイズピークを指す) の相対強度から、30ポイントのうち便宜的に上位5つまでのピークを選択する (表1)。() は演算に用いたEIIP列を示す (以下、同様)。

(表1)

全配列由来の振動数値 (512)	0.4355	0.0645	0.4785	0.4336	0.0664
------------------	--------	--------	--------	--------	--------

第二段階は、それら5つのピークから活性部位 (マゲイニン2) に由来するピークを選択するために、前駆体の全アミノ酸周波数スペクトルと活性部位領域と仮定したマゲイニン2のアミノ酸配列から活性部位周波数スペクトルを求め、それらをクロスした (図3)。図3から活性部位由来のピークを表2にまとめた。

(表2)

活性部位由来の振動数値 (512) 0. 4355 0. 4336 0. 0645 0. 0664 0. 2598

第三段階は、第二段階で得られた5つのピークが活性部位に由来しているかどうか確認するために、前駆体と活性部位領域のアミノ酸残基を全てロイシンで置換した置換全アミノ酸周波数スペクトルと全アミノ酸周波数スペクトルをクロスし、図4のような前駆体の全アミノ酸周波数スペクトルが得られた。図2や図3と同じ条件でピークを選択すると表3になる。但し、この場合、置換に用いるアミノ酸残基はロイシン以外の19種類のアミノ酸の何れでも構わない。

(表3)

置換全配列由来の振動数値 (512) 0. 4355 0. 0645 0. 4785 0. 4336 0. 0664

表1、表2、表3を整理すると表4になる。

(表4)

マゲイニン

全配列由来の振動数値 (512) 0. 4355 0. 0645 0. 4785 0. 4336 0. 0664

活性部位由来の振動数値 (512) 0. 4355 0. 4336 0. 0645 0. 0664 0. 2598

置換全配列の振動数値 (512) 0. 4355 0. 0645 0. 4785 0. 4336 0. 0664

表4から判るように、0. 0664、0. 2598及び0. 4336のピークが活性部位由来の優先的なピークである。

更に、本発明者は鋭意検討した結果、天然型又は非天然型の任意のアミノ酸配列(図1B)とそのアミノ酸配列に学術的に対応する核酸配列(図1E)からDFTによって得られる全アミノ酸周波数スペクトル(図1D)と全核酸周波数スペクトル(図1F)をクロス

することにより混合周波数スペクトル（図 1 G）（以下混合周波数スペクトルと称す）をもとめ、効率よく任意の蛋白質の活性部位由来の特異的振動数値を選択し、その蛋白質の生物・機能活性または結合活性を予測する別途方法を見出し、本発明を完成するに到った。

即ち、第一段階はヴェルコヴィッチ等（V. Veljkovic et al, Cancer Biochem. Biophys., 9, 139-148 (1987) ; I. Cosic, IEEE, 41, 1101-1114 (1994)）の方法に従って、任意の全アミノ酸配列のアミノ酸残基に EIIP 指標数を付与し、DFT することによって全アミノ酸周波数スペクトル（図 1 D）を作製する。

アミノ酸残基に与える EIIP 指標値は次の通りである。

Leu 0.0000 Ile 0.0000 Asn 0.0036 Gly 0.0050 Val 0.0057 Glu 0.0058 Pro 0.0198
His 0.0242 Lys 0.0371 Ala 0.0373 Tyr 0.0516 Trp 0.0548 Gln 0.0761 Met 0.0823
Ser 0.0829 Cys 0.0829 Thr 0.0941 Phe 0.0946 Arg 0.0959 Asp 0.1263

第二段階は、第一段階で用いたアミノ酸配列に学術的に対応する核酸配列の核酸残基に EIIP 指標数を付与し、第一段階の方法と同様に、その EIIP 列を DFT することに全核酸周波数スペクトル（図 1 F）を得る。核酸残基に与える EIIP 指標値は次の通りである。グアニン: G 0.0806、アデニン: A 0.1260、チミン（ウラシル: T(U) 0.1335 (0.0562)）、シトシン: C 0.1340

第三段階は、第一段階と第二段階で得られた全アミノ酸周波数スペクトルと全核酸周波数スペクトルをクロスして混合周波数スペクトル（図 1 G）を求め、活性部位由来の振動数値を選択する。

即ち、マゲイニン 2 前駆体のアミノ酸配列とその核酸配列から混合周波数スペクトルを得る（図 5）。図 5 から、ピークの相対強度に従うと 0.2607、0.4346、0.4785、0.3916、0.0215、0.0654 等の優先的な振動数値が選択される。同様にして、活性部位領域と仮定したマゲイニン 2 の混合周波数スペクトルを求めると図 6 になる。図 6 から、ピークの相対強度に従うと 0.2656、0.0625、0.2422、0.2500、0.0547、などの優先的な振動数値が選択される。図 5 と図 6 のスペクトルの比較から、前記 6 つの振動数値の 0.0743 等うち 0.2607 と 0.0654 がマゲイニン 2 領域由来のピークであることが判る。又、それらの数値は表 4 の 0.25

98、0.0664の数値に近い。本発明者は、既にマゲイニン2誘導体 (MSI-78A) のオキザロアセテートに対する脱炭酸活性を報告した (N. Numao et al, Biol. Pharm. Bull., 22, 73-76 (1999))。但し、マゲイニン2は脱炭酸活性を殆ど発現しない。MSI-78Aのアミノ酸配列とその学術的に対応する核酸配列を想定し、その混合周波数スペクトル (図7) を求める。図7より、ピークの相対強度に従うと0.1641、0.1719、0.0938、0.2813、0.2422がMSI-78Aの優先的振動数値である。従って、MSI-78Aのオキザロアセテートに対する脱炭酸活性は0.0938、0.1641、0.1719、0.2813に由来するものと思われる。

本発明者は活性部位領域が1箇所だけの参考例として、高カルシウム血症治療薬として知られ、32個のアミノ酸からなる鮭カルシトニン (I) とその前駆体 (136個のアミノ酸) について、活性部位由来の特異的な振動数値を選択した。それらのアミノ酸配列はSWISS-PROTに登録されている。

第一段階として、鮭カルシトニン (I) 前駆体の全アミノ酸配列を自己クロスすることにより、前駆体の全アミノ酸周波数スペクトル (図8) を求め、ピークの相対強度に従って前駆体全分子由来の5つのピークを選択する。第二段階は、活性部位領域として仮定した鮭カルシトニン (I) の活性部位周波数スペクトルと鮭カルシトニン前駆体の全アミノ酸配列を全アミノ酸周波数スペクトルをクロスし、クロススペクトル (図9) から、鮭カルシトニン (I) 由来のピークを選択する。第三段階は、第二段階で得た5つのピークが鮭カルシトニン部位由来のピークかどうかを確認するために、その前駆体中に存在する鮭カルシトニン領域 (アミノ酸番号84~114) をロイシンで置換した置換全アミノ酸周波数スペクトルと全アミノ酸周波数スペクトルをクロスし、得られたクロススペクトル (図10) から、置換全配列の振動数値Iを求める。マゲイニンと同じ条件でピークを選択し、整理すると表5になる。又、第三段階の別法として、前駆体のアミノ酸配列中、鮭カルシトニンがコードされている領域 (アミノ酸番号84~114) のアミノ酸配列以外のアミノ酸残基すべてをロイシンで置換した置換全アミノ酸配列を上法に従って演算した (図11)。その結果を置換全配列の振動数値IIとして表5に記載した。但し、この場合も、置換に用いるアミノ酸残基はロイシン以外の19種類のアミノ酸の何れでも構わない。

(表5)

鯉カルシトニン

全配列由来の振動数値	(256)	0.0469	0.1328	0.1445	0.1992	0.4063
活性部位由来の振動数値	(256)	0.1563	0.2734	0.1445	0.0469	0.1523
置換全配列の振動数値Ⅰ	(256)	0.1250	0.0469	0.0508	0.0195	0.1445
置換全配列の振動数値Ⅱ	(256)	0.0469	0.0195	0.1680	0.0508	0.2734

表5から、0.1445、0.1523、0.1563と0.2734のピークが鯉カルシトニン (I) 由来のピークである。既知文献 (V. Veljkovic et al, Cancer Biochem. Biophys., 9, 139-148 (1987); I. Cosic, IEEE, 41, 1101-1114 (1994)) にはそれらの数値を示す蛋白質が記載されていないので、鯉カルシトニン (I) の新しい生物・機能活性が不明である。然しながら、鯉カルシトニン (I) にはすくなくとも2つ以上の振動数値に由来する生物・機能活性が期待できるだろう。

本発明者は更に、活性部位由来の生物活性が分子生物学的に同定されている参考例として、そのアミノ酸配列のC末端領域が欠失した蛋白質は抗ウィルス活性が殆どなくなることが知られているガンマインターフェロン (R. Wetzel et al, Protein Eng, 3, 611-623 (1987)) について、活性部位由来の特異的な振動数値を選択した。そのアミノ酸配列はSWISS-PROTに登録されており、166個のアミノ酸残基から構成されている。活性部位はガンマインターフェロンのアミノ酸配列中151~154残基に存在することが知られている。併しながら、本発明の目的をより明らかにするために、本発明者が13種類のモチーフを使って予測した活性部位 (247 ± 15) を採用する (N. Numao et al, Biol. Pharm. Bull., 16, 1160-1163 (1993))。

第一段階として166個のアミノ酸残基からなる、ガンマインターフェロンの全アミノ酸配列を自己クロスすることにより、ガンマインターフェロンの全アミノ酸周波数スペクトル (図12) を求める。ピークの相対強度に従って、上位5つ選択する。第二段階は、ガンマインターフェロンの活性部位領域である132~162の活性部位周波数スペクトルとガンマインターフェロンの全アミノ酸配列を全アミノ酸周波数スペクトルをクロスさせ、132~162由来のピークをピークの相対強度に従って、上位5つ選択する (図13)。第三段階は、第二段階で得た5つのピークが132~162由来かどうかを確認するために、ガンマインターフェロンに存在する活性部位領域 (132~162) をロイシンで置換した置換全アミノ酸周波数スペクトルと全アミノ酸周波数スペクトルをクロスし、

図14のような前駆体の全アミノ酸周波数スペクトルが得られた。又、第三段階の別法として、前駆体のアミノ酸配列中、132～162のアミノ酸配列を残し、あとのアミノ酸残基すべてをロイシンで置換した置換全アミノ酸を常法に従って演算した。その結果を置換全配列の振動数値IIとして表6に記載した(0～0.015の数値は明らかに活性部位由来ではないので無視した)。但し、この場合も、置換に用いるアミノ酸残基はロイシン以外の19種類のアミノ酸の何れでも構わない。

(表6)

ガンマインターフェロン

全配列由来の振動数値	(256)	0.3594	0.4023	0.0469	0.1484	0.0781
活性部位由来の振動数値	(256)	0.0234	0.3594	0.3633	0.0273	0.4023
置換全配列の振動数値I	(256)	0.3594	0.0469	0.0781	0.1484	0.0117
置換全配列の振動数値II	(256)	0.0234	0.3594	0.3633	0.4023	0.0273

表6から、0.0234、0.0273、0.3633のピークがガンマインターフェロンの予測活性部位由来のピークである。既知文献(V. Veljkovic et al, Cancer Biochem. Biophys., 9, 139-148 (1987); I. Cosic, IEEE, 41, 1101-1114 (1994))によれば、インターフェロン全分子に由来する振動数値は 0.082 ± 0.008 である。然しながら、本発明の活性部位由来の振動数値はピークの優先性に基づけば、0.0117～0.0234がガンマインターフェロンの抗ウィルス活性に関与しているだろう。既知文献から、0.0234の数値はヘモグロビンの数値と一致する。

更に、本発明者は本発明方法をより明確にするために別途方法を記載する。即ち、ガンマインターフェロンのアミノ酸配列とその学術的に対応する核酸配列から、それらの全周波数スペクトルをクロスして混合周波数スペクトル(図15)を求め、ガンマインターフェロン活性部位由来のピークを選択した。図15から、ピークの相対強度に従って、0.0098、0.1250、0.4043、0.0117、0.334、0.2324等を選択する。これらの数値のうち、0.0010は0.0098とガンマインターフェロン由来の0.0117に近い。更に、本方法のDFT解析を鋭意検討した結果、ガンマインターフェロンの周波数領域($m/L \leq 0.5$)とガンマインターフェロンレセプターの周波数領域($0.5-m/L$)の間に結合に関する一定の規則性を見出し、本発明を完

成した。即ち、ガンマイインター フェロンの受容体であるガンマイインター フェロンレセプター (M. Aguet et al, Cell 55, 273-280 (1988)) の細胞外領域のアミノ酸配列とその学術的に対応する核酸配列から混合周波数スペクトル (図 1 6) を求め、その特異的な振動数を選択した。図 1 6 より、ピークの相対強度に従って、周波数領域 (0.5-m/L) の 0.2412、0.0703、0.3223、0.0010、0.0400、0.4395 等が、その細胞外領域の特異的振動数値として選択された。これらの数値のうち、0.0010、0.2412、0.3223 はガンマイインターフェロン活性部位由来の数値 (0.0098、0.2324、0.3340) に近い。同様の方法でリガンド/レセプター結合関係を説明できる例として、HIVgp120 と CD4、Poliovirus coatprotein VP1 と poliovirus receptor、IL-2 と IL-2 receptor、TNF- α (又は TNF- β) と 55kdTNF receptor または Insulin と Insulin Receptor 等を列挙できる。更に、ガンマイインター フェロンレセプター の優先的ピークの数値は IL-2、TNF- α 、TNF- β 、Insulin 等の優先的振動数値よりもガンマイインターフェロンの数値と多数重複する。従って、本発明方法は任意の蛋白質に対して選択的に結合する別の蛋白質を探索できる。

本発明者は更に、酵母の転写因子蛋白質 Gal4p (A.S. Laughon et al, Mol. Cell Biol., 4, 260-267 (1984)) の活性部位由来の特異的振動数値を新規方法で選択した。Gal4p 蛋白質は 881 個のアミノ酸から構成され、その DNA- 蛋白質結合ドメインは 1 4 ~ 5 7 領域にあることが報告されている (M. Johnston Microbiol. Rev., 51, 458-476 (1987))。即ち、Gal4p のアミノ酸配列とその学術的に対応する核酸配列から得られるそれらの全周波数スペクトルをクロスして混合周波数スペクトル (図 1 7) を求める。図 1 7 から、ピークの相対強度に従って、0.3311、0.2705、0.3818、0.0051、0.3901、0.0796、0.3181、0.1280 等が Gal4p 由来の優先的振動数値として選択される。次に、1 4 ~ 5 7 領域のアミノ酸配列とその学術的に対応する核酸配列から得られるそれらの全周波数スペクトルをクロスして混合周波数スペクトル (図 1 8) を求める。図 1 8 から、ピークの相対強度に従って、0.1289、0.3750、0.0352、0.0391、0.1328、0.2383、0.3789、0.3908 等が 1 4 ~ 5 7 領域由来の優先的振動数値として選択される。従って、図 1 7 と図 1 8 から、Gal4p 由来の優先的振動数値のうち少なくとも 0.1280、0.3818 は DNA- 蛋白質結合に関与している。このことから、Gal4p の全アミノ酸配列周波数スペクトルには活性部位由来の振動数値が含まれている可能性がある。

一方、本発明者は鋭意検討し、任意の核酸配列 (図 1 B) とその一本鎖核酸配列に水

素結合によって形成しうる核酸配列 (図 1 E) から DFT によって得られるそれぞれの全核酸周波数スペクトル (図 1 D と 1 F) をクロスすることにより均一核酸周波数スペクトル (図 1 G) (以下均一核酸周波数スペクトルと称す) をもとめ、得られる優先的な特異的振動数値を選択した。核酸残基に与える EIIP 指標値は、グアニン: G 0.0806、アデニン: A 0.1260、チミン (ウラシル: T (U) 0.1335 (0.0562)、シトシン: C 0.1340 である。即ち、Gal7 プロモーター領域は 350 個の核酸から構成され、酵母の転写因子蛋白質 Gal4p (A.S. Laughon et al, Mol. Cell Biol., 4, 260-267 (1984)) に結合する (R. J. Bram et al, EMBO J., 5, 603-608 (1986))。その一本鎖核酸配列の核酸配列周波数スペクトル (図 1 D) を求める。次に、その核酸配列に対して水素結合によって形成しうる一本鎖核酸配列 (図 1 E) から核酸配列周波数スペクトル (図 1 F) をもとめる。更に、それら 2 つの核酸配列周波数スペクトルをクロスして、均一核酸周波数スペクトル (図 1 F) (図 1 9) を求める。図 1 9 から、ピークの相対強度に従って、周波数領域 (0.5-m/L) の 0.4805、0.0820、0.4210、0.4336、0.1211、0.4844、0.3066、0.2051、0.3867 など Gal7 プロモーター領域の特異的振動数値として選択する。それらの数値のうち、0.0820、0.1211、0.3066、0.3867 は上記記載の Gal4p の、0.0796、0.1280、0.3181、0.3818 とほぼ一致する。従って、Gal4p 由来と Gal7 プロモーター領域由来の優先的な特異的振動数値の重複度合いを考慮すれば、これら 2 つの高分子化合物は結合しうるということが説明できる。このような方法でその結合を合理的に説明できる例として、プリオン蛋白質由来ペプチドと合成 RNA セグメント (S. Weiss et al, J. Virol., 71, 8790-8797 (1997))、OCT1 と TNF プロモーター領域 (J.C. Knight et al, Nature Genetics 22, 145-150 (1999))、pho4p (又は pho2p) と Pho5 転写領域 (Y.Ohsima, Genes Genet. Syst. 72, 323-334 (1997)) 等を列挙できる。従って、本発明方法によって、所望の核酸配列から結合する蛋白質の予測、及び / 又は所望の蛋白質から結合する核酸配列の予測も可能である。

本発明者は、天然型又は非天然型の、1) 任意の蛋白質の全アミノ酸配列の全アミノ酸周波数スペクトルと、その学術的に対応する核酸配列全核酸周波数スペクトルをクロスさせて任意の混合周波数スペクトルを求め、その優先的な特異的振動数値を選択する。2) 別の蛋白質の全アミノ酸配列の全アミノ酸周波数スペクトルと、その学術的に対応する核酸配列の全核酸周波数スペクトルをクロスさせて混合周波数スペクトルを求め、その優先的な特異的振動数値を選択する。次に、それら 2 つの混合周波数スペクトルにおいて、ピークの相対強度に従って選択した優先的な特異的振動数値の重複度を測定し、2 つの蛋白

質の生物・機能活性の類似性を予測する方法を見出し、本発明を完成した。

即ち、セリンプロテアーゼとして分類されているウロキナーゼ (UK) (W. E. Holmes et al, Biochemistry 3, 925-929 (1985)) とサブチリシン (J.A. Wells et al, Nucl. Acids Res., 11, 7911-7925 (1983)) の活性部位由来の特異的振動数値を選択し、それらの生物・機能活性の類似性を検討した。それらのアミノ酸配列は SWISS-PROT に、核酸配列は GenBank 等に登録されている。

即ち、第一段階として、UK の全アミノ酸配列 (431aa) の全アミノ酸周波数スペクトルと、その学術的に対応する核酸配列 (1293na) の全核酸周波数スペクトルを求める。次いで、それらのスペクトルをクロスさせて UK の混合周波数スペクトル (図 20) を求め、ピークの相対強度に従って、その優先的な特異的振動数値を選択する。UK の優先的な特異的振動数値は、0.4136、0.0454、0.0898、0.3608、0.0762、0.0449、0.4141、0.4814、0.2061、0.4009 等となる。第二段階として、サブチリシンの全アミノ酸配列 (376aa) の全アミノ酸周波数スペクトルと、その学術的に対応する核酸配列 (1128na) の全核酸周波数スペクトルを求める。次いで、それらのスペクトルをクロスさせてサブチリシンの混合周波数スペクトル (図 21) を求め、ピークの相対強度に従って、その優先的な特異的振動数値を選択する。サブチリシンの優先的な特異的振動数値は、0.3330、0.3335、0.3169、0.1973、0.0415、0.3325、0.2397、0.2412、0.2075、0.1191 等となる。図 20 と図 21 から、優先的な特異的振動数値から 3 つ以上の重複ピーク (0.0454, 0.0449, 0.2061 と 0.0415, 0.1973, 0.2075) が選択できる。このような方法によって、生物・機能活性の類似性を説明できる典型的な例として、TNF- α と TNF- β 、上記に述べた酵母の転写因子である pho4p と pho2p 等が列挙できる。従って、本発明方法によって、所望の蛋白質と類似の生物・機能活性を有する蛋白質の予測も可能である。

本発明者は、更にプロモター領域の核酸配列の全核酸配列周波数スペクトルと、その学術的に対応する核酸配列全核酸周波数スペクトルをクロスさせて均一核酸周波数スペクトルを求め、ピークの相対強度に従って、その優先的な特異的振動数値を選択すると、そこにはモチーフ由来の優先的な特異的な振動数値が含まれていることを明らかにしている。更に、この応用例として、ゲノム配列上のエキソンとイントロンとの相互作用にも利用できる。このような例として、CD4 等の未成熟 mRNA 上にあるエキソンとイントロンとの相

相互作用を例示できる。

以上まとめると、本発明者は GenBank、EMBL、PIR、SWISS-PROT 等のデータベースに登録された機能未知の天然型、非天然型由来の、1) 任意のアミノ酸配列のアミノ酸残基に、EIIP (Electron-ion interaction potential) 指標数を付与し、得られた EIIP 列を DFT して得られる全アミノ酸配列周波数スペクトル及び/又は 2) 任意のアミノ酸配列に存在し、活性部位に関与する既知モチーフを 1 つ以上含有する 2 ～ 64 個のアミノ酸残基を含有するアミノ酸配列領域のアミノ酸に EIIP 指標数を付与し、得られた EIIP 列を DFT して得られる活性部位周波数スペクトル及び/又は 3) 任意のアミノ酸配列のアミノ酸配列に学術的に対応する核酸配列領域の核酸残基に EIIP 指標数を付与し、得られた EIIP 列を DFT して得られる全核酸配列周波数スペクトル及び/又は 4) 任意の核酸配列の核酸残基に EIIP 指標数を付与し、得られた EIIP 列を DFT して得られる全核酸配列周波数スペクトル及び/又は 5) 任意の核酸配列に水素結合する核酸配列の核酸残基に EIIP 指標数を付与し、得られた EIIP 列を DFT して得られる全核酸配列周波数スペクトルを、組み合わせて任意のアミノ酸配列 (又は核酸配列) の生物・機能活性 (又は結合活性) を予測する方法を見出し、本発明を完成するに至った。

従って、本発明方法の作製過程より明らかなことは、前記の概念を数学的手法 (フーリエ解析、ウェーブレット解析等) 用いて作製したプログラムや記憶媒体によって予測した、天然型乃至非天然型の任意の蛋白質 (又は核酸配列) の整理、分類した一覧表及び/又は機能・活性、結合相関図は決して本発明の例示のみに限定されるわけではないことも明らかである。従って、前記概念によって作成された一覧表から、任意の蛋白質や核酸配列と所望の相互作用が期待できるような新しい蛋白質や核酸配列を探すことも容易に推測できるので、そのようなプログラム開発や記憶媒体も本発明の基本概念を使う限り本請求範囲であることも明らかである。又、前記のアミノ酸残基と核酸残基の EIIP 指標数の代替として、それらの親水性や疎水性の定数や単なる整数などを用いて、前記の概念を数学的手法 (フーリエ解析、ウェーブレット解析等) 用いて作製したプログラムや記憶媒体によって予測した、天然型乃至非天然型の任意の蛋白質 (又は核酸配列) の整理、分類した一覧表及び/又は機能・活性、結合相関図も本請求範囲である。

更に、マゲイニン前駆体やガンマインターフェロン、Gal4p の混合周波数スペクトルで得られた優先的な特異的振動数値に着目すれば任意の配列（アミノ酸配列、核酸配列）の活性部位を絞り込めるので、本発明方法を活性部位領域からの手段のみに限定している訳ではない。更に、そのような任意のアミノ酸配列又は核酸配列の活性部位の絞込みに関するプログラム開発や記憶媒体も当然本請求項の範囲である。

更に、アミノ酸配列の方向性を N→C から C→N に変えても、核酸配列の方向性を 5' → 3' を 3' → 5' に変えても本方法で得られる結果は全く同じ結果が得られることが既に、判明しているので、それらの配列方向性が限定されている訳ではない。

更に、本発明方法はアミノ酸配列又は核酸配列中に存在する活性部位に関する根源的な原理又は概念の探究に基づいているので、予測した蛋白質又は核酸配列の機能・活性の用途の請求範囲は、単に治療薬としての農薬や医薬品等だけでなく、遺伝病や疫病などの予防・診断法も本請求範囲であることは明白である。

以下、本発明を実施例により具体的に説明するが本発明はこれらに限定されるものではない。

（実施例 1） 正常プリオン蛋白質の生物・機能活性の予測

正常プリオン蛋白質の生物活性は現在まで報告されていない（S.B. Prusiner, Proc. Natl. Acad. Sci., U.S.A., 95, 13363-13383 (1997); D. Westway et al., Proc. Natl. Acad. Sci., U.S.A., 95, 11030-11031 (1998)）。そのアミノ酸配列は既に SWISS-PROT に登録されている。既知文献（G. Forloni et al, Nature, 362, 543-546 (1993)）によれば、プリオン蛋白質のアミノ酸番号 106～126 付近に神経毒性活性があることが知られている。併しながら、その配列ペプチドを緩衝溶液（pH7.4）中、37℃30 日間処理すると、そのペプチドは神経毒性活性を発現しないとの反証もある（B. Kunz et al, FEBS Lett., 458, 65-68 (1999)）。本発明者は、まずプリオン全アミノ酸周波数スペクトルの自己クロス、プリオン全アミノ酸周波数スペクトルと活性部位（アミノ酸番号 109～131）の周波数スペクトルとのクロススペクトルを求める。次に、プリオン全アミノ酸周波数スペクトルと活性部位領域のアミノ酸残基をロイシンで置換した置換プリオン全アミノ酸周波数スペクトルとのクロスを行った（図 22, 23, 24）。結果を表 7 に示す。

(表 7)

プリオン

全配列由来の振動数値	(256)	0.0039	0.2617	0.3164	0.4961	0.3789
活性部位由来の振動数値	(256)	0.2617	0.2539	0.3164	0.0234	0.0313
置換全配列由来の振動数値 I	(256)	0.0039	0.2617	0.4961	0.3614	0.1367

表 7 から、0.2617、0.2539、0.0234、0.0313 などの特異的振動数値が 109～131 領域由来のピークである。これらの数値はプリオン蛋白質の全アミノ酸配列周波数スペクトルとマゲイニン 2 誘導体 (MSI-78A) の全アミノ酸周波数スペクトルとのクロススペクトルでも保存されている。このことから 109～131 領域は MSI-78A と類似の生物・機能活性が期待できる。更に、その予測を検討するために、プリオン蛋白質の 109～131 領域の混合スペクトル (図 25) と MSI-78A の混合スペクトル (図 7) を比較した。図 25 より、ピークの相対強度に従って優先的振動数値として選択された 0.0391、0.0547、0.0313、0.3203、0.2969、0.1016、0.0938 等のうち、0.1016 と 0.0938 が MSI-78A の 0.0938 と一致ないし近似値となった。その 2 つの振動数値のうち、0.0938 はプリオン蛋白質の 106～126 領域の混合スペクトル (図 26) では最優先振動数値となることが示唆された。これらの結果より、109～131 領域のアミノ酸配列中に存在する 106～126 領域ばかりでなく、正常なプリオン蛋白質分子全体の生物活性の 1 つとして脱炭酸活性が期待できるだろう。更に、既知文献 (I. Cosic, IEEE, 41, 1101-1114 (1994)) に基づくと、図 26 で得られる振動数値 0.0625 や 0.0781 はミオグロビンやチトクロームなどの振動数値に近い。事実、プリオン蛋白質は銅結合蛋白質であり抗酸化剤として生体反応に関与していることが報告されている (D.R. Brown et al, J. Neurochem., 76, 69-76 (2001))。

(実施例 2) アミロイド蛋白質前駆体 (APP) の機能予測

3 種類の APP のうち、1 つは 751 個のアミノ酸残基からなる蛋白質である (A. Ponte et al., Nature, 331, 525-527 (1988))。そのアミノ酸配列は既に SWISS-PROT に登録されている。APP の機能活性部位は蛋白質機能部位予測法 (N. Numao et al, Biol. Pharm. Bull., 16, 1160-1163 (1993)) においてアミノ酸番号 142、340、513、665 付近に活

性部位が存在することが既に予測されている。これらのうち、特にモチーフ (VXH, KL, GA) を含む領域 (アミノ酸番号650～680) の生物活性は老人性痴呆症に絡んでいることが報告されている。本発明者はその領域の新しい生物・機能活性を探し出すために本方法を検討した。即ち、その領域を活性部位とし、上記の方法に従って検討した (図27、28、29)。結果を表8に示す。

(表8)

APP

全配列由来の振動数値	(1024)	0.4277	0.3818	0.3701	0.0283
		0.3610			
活性部位由来の振動数値	(1024)	0.3203	0.2588	0.3701	0.3818
		0.3193			
置換全配列由来の振動数値I	(1024)	0.4277	0.3818	0.0361	0.3701
		0.0283			

表8から、APPの活性部位由来の振動数値は0.3193～0.3203と0.2588であることが示唆された。0.3193～0.3203の数値は、既知文献に基づくグルカゴン (0.3203 ± 0.034) とリゾチーム (0.3281 ± 0.004) に近い。

(実施例3) APP中にあるプロテアーゼ阻害活性領域の活性予測

APPのアミノ酸番号で291～341付近のアミノ酸配列はセリンプロテアーゼ阻害剤と相同性が高いということも既に知られている (A. Ponte et al., Nature, 331, 525-527 (1988))。事実、この領域の阻害活性が既に報告されている (N. Kitaguchi et al., Nature, 331, 530-532 (1988)) が、その阻害活性は高くない。その実験結果を参考にして、アミノ酸番号289～364を活性部位領域として本方法を適用し、演算を試みた (図30)。その結果を表9に示す。

(表9)

APP 中の289～364領域における活性予測

全配列由来の振動数値	(1024)	0.4277	0.3818	0.3701	0.0283
		0.3610			
活性部位由来の振動数値	(1024)	0.3818	0.3203	0.2587	0.4277
		0.3701			

既知文献 (I. Cosic, IEEE, 41, 1101-1114 (1994)) に従えば、プロテアーゼ阻害剤の振動数値は 0.3555 ± 0.008 なので、上記の上位2つの数値は一致しない。因みに、既知文献 (A. Ponte et al., Nature, 331, 525-527 (1988)) に記載された Kunitz protease inhibitors を演算したところ、その特異的な振動数値は 0.3281 であった。上記の5つの数値のうち、 0.3203 がそれに近い。従って、実施例2のAPPの650～680領域の新しい生物活性として阻害活性が期待できるかもしれない。

(実施例4) ヒト成長ホルモン (hGH) の新しい生物活性予測

蛋白質合成、軟骨発育促進及び脂肪分解作用を有する217個のアミノ酸からなるヒト成長ホルモン (hGH) について、活性部位由来の新しい生物・機能活性が期待できるかを本方法で検討した。そのアミノ酸配列は既に SWISS-PROT に登録されている。その活性部位はアミノ酸番号205付近に活性部位が存在することが予測されている (N. Numao et al, Biol. Pharm. Bull., 16, 1160-1163 (1993)) ので、それを参考にし、本方法の適用法は実施例2と同様に処理した。即ち、アミノ酸番号197～217領域を活性部位として検討した (図31, 32, 33)。結果を表10に示す。

(表10)

ヒト成長ホルモン

全配列由来の振動数値	(256)	0.1328	0.4570	0.4336	0.1719	0.3945
置換全配列の振動数値 I	(256)	0.1328	0.0234	0.2578	0.4336	0.4258
置換全配列の振動数値 II	(256)	0.0234	0.1719	0.1641	0.1680	0.0508

表10から、hGHの活性部位依存の生物活性は 0.0234 と $0.1641 \sim 0.17$

19の振動数値であるが、前者は活性部位領域以外の振動数値と重複する。このことから、活性は分子全体に由来する。事実、hGHには活性部位が3箇所存在することが既に知られている (B.C. Cunningham et al., Science, 244, 1081-1085 (1989))。然しながら、0.1641~0.1719の振動数値は鮭カルシトニンの活性部位由来の振動数値 (0.1445~0.1563) の近傍である。そこで、hGHの全アミノ酸周波数スペクトルと鮭カルシトニン前駆体周波数スペクトル又は32個のアミノ酸からなる鮭カルシトニン周波数スペクトルとクロスした (図34、図35)。その結果、0.1328~0.1719のピークの優先性が観察された。このことから、hGHと鮭カルシトニンには類似の生物活性が期待できる。

(実施例5) データベースの構築

蛋白質の活性部位に由来する生物機能に関するデータベースの構築法は色々考えられるが、本発明を見いだした過程や実施例で得られた結果 (表1~10) を整理し、それらをコンピュータを使って検索し易いように構築すること。表11に一例を示す。

(表11)

マゲイニン

活性 0.4355 0.4336 0.0645 0.0664 0.2598

鮭カルシトニン

活性 0.1563 0.2734 0.1445 0.0469 0.1523

ガンマイインターフェロン

活性 0.0234 0.3594 0.3633 0.0273 0.4023

APP

活性(アミロイド領域) 0.3203 0.2588 0.3701 0.3818 0.3193

活性(阻害領域) 0.3818 0.3203 0.2587 0.4277 0.3701

ヒト成長ホルモン

活性 0.0234 0.1718 0.1641 0.1680 0.0508

こうして本方法により種々の蛋白質の演算を行い、別途に活性評価の結果を加えていけば従来の蛋白質の機能予測より極めて優れた有用データベースが作製できることは容易に類推できる。従って、本発明記載以外の他の蛋白質へ本発明の基本概念を利用する限り、その結果に基づいたデータベースの構築は本請求の範囲にある。

(実施例6) データベースの活用

表11のデータベースから、APPにも脱炭酸活性が期待できるかもしれない。APPの650～680領域とプリオン蛋白質にはプロテアーゼ阻害活性が期待できる。カルシトニンとhGHには類似の生物活性が期待できる。hGHとガンマイインターフェロンには類似の活性が期待できる。

(実施例7) エボラウイルス結合蛋白の予測

エボラウイルスはウイルス性出血熱を発症させる国際伝染病の1つとして知られている。

併しながら、依然としてそのレセプター蛋白質が報告されていない。本発明で使った幾つかのレセプター蛋白質とエボラウィルスのエンベロープ蛋白(図36)(V.E. Volchkov et al, Virology 214, 421-430 (1995))との結合活性予測を行ったところ、CD4、poliovirus receptor、IL-2 receptor、55kdTNF receptor、Insulin Receptorの中で、55kdTNF receptorの細胞外領域との相互作用が他のレセプターよりも高く予測された。

本方法は従来のランダム法に近い活性評価法よりも、予め既知の蛋白質と核酸配列データベースに基づいて、任意の蛋白質や核酸配列の新しい機能活性又は結合相手を予測できるので、極めて有用である。

【特許請求の範囲】 (What is claimed is:)

1. 天然型、非天然型由来の任意のアミノ酸配列のアミノ酸残基に、EIIP (Electron-ion interaction potential) 指標数を付与し、得られた数値列 (EIIP 列) を離散フーリエ変換 (DFT) して得られた全アミノ酸配列周波数スペクトルと、

前記任意のアミノ酸配列に存在し、活性部位に関与する既知モチーフを1つ以上含有する2～64個のアミノ酸残基を含有するアミノ酸配列領域のアミノ酸にEIIP 指標数を付与し、得られたEIIP 列をDFT して得られた活性部位周波数スペクトルとを求め、

前記全アミノ酸配列周波数スペクトルと前記活性部位周波数スペクトルとをクロスしたスペクトルから、その蛋白質の活性部位由来の特異的振動数値を選択し、特異的振動数値の近似値を検索することにより任意の蛋白質の生物・機能活性及び/又は結合活性を予測する方法。

2. 活性部位の信号としての既知モチーフとして、GT、AS、GA、ID、TR、SR、LK、TXW、VXH、MXH、WXP、AXC、GXS (G、T、A、S、I、D、R、L、K、W、V、H、M、P、C、Xはグリシン、スレオニン、アラニン、セリン、イソロイシン、アスパラギン酸、アルギニン、ロイシン、リジン、トリプトファン、バリン、ヒスチジン、メチオニン、プロリン、システイン及び20種類のアミノ酸の何れかを意味する) 及び/又はそれらの逆配列の何れか1つ以上を用いる請求項1に記載の任意の蛋白質の生物・機能活性及び/又は結合活性を予測する方法。

3. 天然型、非天然型由来の任意のアミノ酸配列のアミノ酸残基に、EIIP (Electron-ion interaction potential) 指標数を付与し、得られた数値列 (EIIP 列) を離散フーリエ変換 (DFT) して得られた全アミノ酸配列周波数スペクトルと、

前記アミノ酸配列に学術的に対応する核酸配列領域の核酸残基にEIIP 指標数を付与し、得られたEIIP 列をDFT して得られた全核酸配列周波数スペクトルとを求め、

前記全アミノ酸配列周波数スペクトルと前記全核酸配列周波数スペクトルとをクロスしたスペクトルからその蛋白質の活性部位由来の特異的振動数値を選択し、特異的振動数値の近似値を検索することにより任意の蛋白質の生物・機能活性及び/又は結合活性を予測す

る方法。

4. 天然型、非天然型由来の任意の一本鎖の核酸配列の核酸残基に EIIP (Electron-ion interaction potential) 指標数を付与し、得られた数値列 (EIIP 列) を離散フーリエ変換 (DFT) して得られた第 1 の全核酸配列周波数スペクトルと、

前記核酸配列に水素結合によって結合する核酸配列の核酸残基に EIIP 指標数を付与し、得られた EIIP 列を DFT して得られた第 2 の全核酸配列周波数スペクトルとを求め、

前記第 1 の全核酸配列周波数スペクトルと前記第 2 の全核酸配列周波数スペクトルとをクロスしたスペクトルからその核酸配列の特異的振動数値を選択し、特異的振動数値の近似値を検索することにより任意の核酸配列の生物・機能活性及び/又は結合活性を予測する方法。

5. 天然型、非天然型由来の任意のアミノ酸配列のアミノ酸残基に、EIIP (Electron-ion interaction potential) 指標数を付与し、得られた数値列 (EIIP 列) を離散フーリエ変換 (DFT) して得られた全アミノ酸配列周波数スペクトルである第 1 のスペクトルと、

任意のアミノ酸配列に存在し、活性部位に関与する既知モチーフを 1 つ以上含有する 2 ～ 64 個のアミノ酸残基を含有するアミノ酸配列領域のアミノ酸に EIIP 指標数を付与し、得られた EIIP 列を DFT して得られた活性部位周波数スペクトルである第 2 のスペクトルと、

アミノ酸配列に学術的に対応する核酸配列領域の核酸残基に EIIP 指標数を付与し、得られた EIIP 列を DFT して得られた全核酸配列周波数スペクトルである第 3 のスペクトルと、

天然型、非天然型由来の任意の一本鎖の核酸配列の核酸残基に EIIP 指標数を付与し、得られた EIIP 列を DFT して得られた全核酸配列周波数スペクトルである第 4 のスペクトルと、

核酸配列に水素結合によって結合する核酸配列の核酸残基に EIIP 指標数を付与し、得られた EIIP 列を DFT して得られた全核酸配列周波数スペクトルである第 5 のスペクトルと

のうち、少なくとも 2 つ以上のスペクトルを求め、求めたスペクトルを組み合わせることによって、天然型、非天然型由来の任意のアミノ酸配列と別の核酸配列との生物・機能活性及び/又は結合活性を予測する方法。

6. 活性部位の信号としての既知モチーフとして、GT、AS、GA、ID、TR、SR、LK、TXW、VXH、MXH、WXP、AXC、GXS (G、T、A、S、I、D、R、L、K、W、V、H、M、P、C、Xはグリシン、スレオニン、アラニン、セリン、イソロイシン、アスパラギン酸、アルギニン、ロイシン、リジン、トリプトファン、バリン、ヒスチジン、メチオニン、プロリン、システイン及び20種類のアミノ酸の何れかを意味する) 及び/又はそれらの逆配列の何れか1つ以上を用いる請求項5に記載の任意のアミノ酸配列と別の核酸配列との生物・機能活性及び/又は結合活性を予測する方法。

7. 天然型、非天然型由来の任意のアミノ酸配列のアミノ酸残基に、EIIP (Electron-ion interaction potential) 指標数を付与し、得られた数値列 (EIIP 列) を離散フーリエ変換 (DFT) して得られた全アミノ酸配列周波数スペクトルである第1のスペクトルと、

任意のアミノ酸配列に存在し、活性部位に関与する既知モチーフを1つ以上含有する2～64個のアミノ酸残基を含有するアミノ酸配列領域のアミノ酸にEIIP 指標数を付与し、得られたEIIP 列をDFT して得られた活性部位周波数スペクトルである第2のスペクトルと、

アミノ酸配列に学術的に対応する核酸配列領域の核酸残基にEIIP 指標数を付与し、得られたEIIP 列をDFT して得られた全核酸配列周波数スペクトルである第3のスペクトルと、

天然型、非天然型由来の任意の一本鎖の核酸配列の核酸残基にEIIP 指標数を付与し、得られたEIIP 列をDFT して得られた全核酸配列周波数スペクトルである第4のスペクトルと、

核酸配列に水素結合によって結合する核酸配列の核酸残基にEIIP 指標数を付与し、得られたEIIP 列をDFT して得られた全核酸配列周波数スペクトルである第5のスペクトルと

のうち、少なくとも2つ以上のスペクトルを求め、求めたスペクトルを組み合わせることによって、天然型、非天然型由来の任意のアミノ酸配列又は核酸配列の活性部位を予測する方法。

8. 活性部位の信号としての既知モチーフとして、GT、AS、GA、ID、TR、SR、LK、TXW、VXH、MXH、WXP、AXC、GXS (G、T、A、S、I、D、R、L、K、W、V、H、M、P、C、Xはグリシン、スレオニン、アラニン、セリン、イソロイシン、アスパラギン酸、アルギニン、ロイシン、リジン、トリプトファン、バリン、ヒスチジン、メチオニン、プロリン、システイ

ン及び20種類のアミノ酸の何れかを意味する) 及び/又はそれらの逆配列の何れか1つ以上を用いる請求項7に記載の任意のアミノ酸配列又は核酸配列の活性部位を予測する方法。

9. 天然型、非天然型由来の任意のアミノ酸配列のアミノ酸残基に、EIIP (Electron-ion interaction potential) 指標数を付与し、得られた数値列 (EIIP 列) を離散フーリエ変換 (DFT) して得られた全アミノ酸配列周波数スペクトルである第1のスペクトルと、

任意のアミノ酸配列に存在し、活性部位に関与する既知モチーフを1つ以上含有する2～64個のアミノ酸残基を含有するアミノ酸配列領域のアミノ酸にEIIP 指標数を付与し、得られたEIIP 列をDFT して得られた活性部位周波数スペクトルである第2のスペクトルと、

アミノ酸配列に学術的に対応する核酸配列領域の核酸残基にEIIP 指標数を付与し、得られたEIIP 列をDFT して得られた全核酸配列周波数スペクトルである第3のスペクトルと、

天然型、非天然型由来の任意の一本鎖の核酸配列の核酸残基にEIIP 指標数を付与し、得られたEIIP 列をDFT して得られた全核酸配列周波数スペクトルである第4のスペクトルと、

核酸配列に水素結合によって結合する核酸配列の核酸残基にEIIP 指標数を付与し、得られたEIIP 列をDFT して得られた全核酸配列周波数スペクトルである第5のスペクトルと

のうち、少なくとも2つ以上のスペクトルを求め、求めたスペクトルを組み合わせることによって、任意のアミノ酸配列及び/又は任意の核酸配列の生物・機能活性及び/又は結合活性を予測する方法。

10. 活性部位の信号としての既知モチーフとして、GT、AS、GA、ID、TR、SR、LK、TXW、VXH、MXH、WXP、AXC、GXS (G、T、A、S、I、D、R、L、K、W、V、H、M、P、C、Xはグリシン、スレオニン、アラニン、セリン、イソロイシン、アスパラギン酸、アルギニン、ロイシン、リジン、トリプトファン、バリン、ヒスチジン、メチオニン、プロリン、システイン及び20種類のアミノ酸の何れかを意味する) 及び/又はそれらの逆配列の何れか1つ以上を用いる請求項9に記載の任意のアミノ酸配列及び/又は任意の核酸配列の生物・機能活性及び/又は結合活性を予測する方法。

11. 請求項1乃至10のいずれか1項の方法を数学的手段を用いて作製し、所望の蛋白質、アミノ酸配列、核酸配列の新しい生物・機能及び/又は結合活性を予測できるような

機能をコンピュータに実現させることを特徴とするプログラム。

12. 前記数学的手段がフーリエ解析又はウェーブレット解析であることを特徴とする請求項11に記載のプログラム。

13. 請求項1乃至10のいずれか1項の方法を数学的手段を用いて作製し、所望の蛋白質、アミノ酸配列、核酸配列の新しい生物・機能及び/又は結合活性を予測できるような機能をコンピュータに実現させるプログラムを記憶したコンピュータ読み取り可能な記憶媒体。

14. 請求項1乃至10のいずれか1項の記載の方法によって予測された少なくとも2種類以上の任意の蛋白質（又はアミノ酸配列、核酸配列）の結合様式。

15. 請求項1乃至10のいずれか1項の記載の方法によって予測された任意の蛋白質又は核酸配列の生物・機能活性の用途であって、

治療薬としての農薬、治療薬としての医薬、遺伝病の予防、遺伝病の診断、疫病の予防、及び疫病の診断のうちから選ばれた少なくとも1つに用いる。

【開示の要約】 (Abstract of the Disclosure)

天然型、非天然型の任意の蛋白質の全アミノ酸配列又は核酸配列に、任意のアミノ酸配列又は核酸配列のアミノ酸（又は核酸）残基に EIIP 指標数を付与し、得られた EIIP 列を DFT して得られた周波数スペクトルを組み合わせ、任意のアミノ酸配列（又は核酸配列）の生物・機能活性又は結合相手を効率よくを予測する。